



HAL
open science

Speeding Up Discovery of Auxetic Zeolite Frameworks by Machine Learning

Romain Gaillac, Siwar Chibani, François-Xavier Coudert

► **To cite this version:**

Romain Gaillac, Siwar Chibani, François-Xavier Coudert. Speeding Up Discovery of Auxetic Zeolite Frameworks by Machine Learning. *Chemistry of Materials*, 2020, 32 (6), pp.2653-2663. 10.1021/acs.chemmater.0c00434 . hal-02518139

HAL Id: hal-02518139

<https://hal.science/hal-02518139>

Submitted on 25 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Speeding Up Discovery of Auxetic Zeolite Frameworks by Machine Learning

Romain Gaillac,^{†,‡} Siwar Chibani,[†] and François-Xavier Coudert^{*,†}

[†]*Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, 75005 Paris, France*

[‡]*Centre de Recherche Paris Saclay, Air Liquide, 78354 Jouy-en-Josas, France*

E-mail: fx.coudert@chimieparistech.psl.eu

Abstract

The characterization of the mechanical properties of crystalline materials is nowadays considered a routine computational task in DFT calculations. However, its high computational cost still prevents it from being used in high-throughput screening methodologies, where a cheaper estimate of the elastic properties of a material is required. In this work, we have investigated the accuracy of force field calculations for the prediction of mechanical properties, and in particular for the characterization of the directional Poisson’s ratio. We analyze the behavior of about 600,000 hypothetical zeolitic structures at the classical level (a scale three orders of magnitude larger than previous studies), to highlight generic trends between mechanical properties and energetic stability. By comparing these results with DFT calculations on 991 zeolitic frameworks, we highlight the limitations of force field predictions, in particular for predicting auxeticity. We then used this reference DFT data as a training set for a machine learning algorithm, showing that it offers a way to build fast and reliable predictive models for anisotropic properties. The accuracies obtained are, in particular, much better than the current “cheap” approach for screening, which is the use of force

fields. These results are a significant improvement over the previous work, due to the more difficult nature of the properties studied, namely the anisotropic elastic response. It is also the first time such a large training data set is used for zeolitic materials.

Introduction

Zeolites are natural and artificial porous aluminosilicates, known since 1756 and artificially synthesized since the 1940s, with generic chemical formula $M_{x/m}Al_xSi_{1-x}O_2$. This family of nanoporous materials has been thoroughly studied for their structural, physical and chemical properties, in particular as they are linked to numerous industrial applications in the fields of adsorption and catalysis. For example, the widely-used zeolite A (also called “LTA”) substituted with different extra-framework cations is used to remove carbon dioxide from natural gas, as dehydrating agent, or to selectively extract linear alkanes in refining processes. Although zeolites have been widely studied by the research community, with our understanding of many of their properties rapidly advancing, the question of the synthetic feasibility of zeolitic frameworks remains a conundrum. While it is mathematically possible to create an infinite number of zeolitic structures, i.e., assemblies of tetrahedra linked by their corners forming periodic frameworks,¹ only a limited number of such frameworks (237 to date²) have been experimentally identified. Even when considering the energetic aspects, by looking at the enthalpy of formation of zeolitic frameworks, the number of “feasible” frameworks remains of the order of hundreds of thousands, and up to few millions, depending on the studies.^{3–5} A new aspect of this question of feasibility of zeolitic frameworks was advanced recently, hinting that mechanical stability — among other physical properties — could play a key role on the determination of feasible zeolite structures.⁶

For this and other reasons, there has been a growing interest in the determination — both experimental and computational — of the detailed mechanical properties of zeolites. The first such theoretical study, to our knowledge, was reported by Astala *et al.* in 2004:⁷ the authors investigated the mechanical properties of 5 different pure silica zeolites via density functional theory (DFT) calculations at the local density approximation (LDA) level. Due to the small number of materials studied, however, this could not lead to useful trends or systematic structure–properties relations for zeolitic frameworks in general. In most subsequent studies, mechanical properties were reported as an aside, when a research group studies a given

framework for a specific application. For instance, Li *et al.* in 2006⁸ looked at low dielectric constant materials among pure silica zeolites for applications in microprocessors. Our group’s prior work, in 2013, was the first systematic study of the elastic properties of known zeolites at the quantum chemistry level.⁶ From conclusions based on the elastic properties of 121 pure silica zeolites, it suggested a feasibility criterion for zeolites based on their elastic anisotropy, as well as the fact that a small number of hypothetical pure silica zeolites could behave as mechanical metamaterials.^{9–11}

In general, the computational determination of the mechanical properties can be performed either at the force field level,¹² or using so-called first principle methods (i.e., at the quantum chemical level).¹³ The differences between the two techniques have to do with transferability, accuracy, and computational cost. In 2015, Jong *et al.* have calculated at the density functional theory (DFT) level the elastic information for 1,181 inorganic compounds,¹³ and all the resulting data are stored on the largest database of calculated elastic properties named Materials Project (MP). At the time of writing, the MP database has been extended and now contains elastic information for 13,751 inorganic compounds. Therefore, despite advances in the computational power of high-performance computing (HPC) resources, DFT calculations of elastic properties are still demanding enough that they have only been performed for a small fraction of all known crystalline materials — or materials whose structure has been studied by DFT.

Computational techniques based on machine learning offer an alternative pathway to discover new materials with targeted mechanical responses, with a much lower computational cost.^{14,15} Machine learning has been used in many areas of chemistry, chemical engineering and materials science, including for the hydrogen storage performance,¹⁶ xenon/krypton separation¹⁷ and prediction of properties (solubility and crystallization).¹⁸ The application of machine learning has also successfully been used in i) micro structural characterization,¹⁹ ii) catalyst development for greenhouse gas conversion, iii) materials discovery for energy harvesting and storage,²⁰ iv) the prediction of magnetic and optoelectronic materials.²¹

When it comes to mechanical properties, in 2016, Jong *et al.* have used machine learning with GBM-Locfit (multivariate local regression within a gradient boosting framework) to predict bulk and shear moduli (K and G , respectively) for inorganic compounds.²² Based on this work, K and G predictions were made available for the whole Materials Project database, for materials where the elastic tensors have not yet been calculated by DFT. Evans *et al.* used a machine learning approach, relying purely on geometric features, to predict the elastic response of 590,448 hypothetical zeolite.²³ The authors combined the gradient boosting regressor (GBR) model using regression trees and a set of local, structural and porosity descriptors on an entire database of hypothetical pure-silica zeolite materials.

Due to the complex structure of zeolites, some of these porous aluminosilicates display unusual mechanical properties, called anomalous properties, such as negative linear compressibility (when a material expands in one specific direction when submitted to an isotropic hydrostatic pressure) or auxeticity (linear expansion in one direction upon elongation in a transverse direction; that corresponds to a negative Poisson’s ratio). These properties have been proposed for several applications: to make sensors and actuators, store mechanical energy, develop new materials with targeted mechanical responses, etc.¹⁰ In the area of negative linear compressibility (NLC), relatively few works have discussed its possible occurrence in zeolitic materials.²⁴ However, a recent systematic study of inorganic materials has shown that it is more frequent than usually estimated,²⁵ and it is not as rare as a property as negative Poisson’s ratio, or auxeticity.²⁶ Auxeticity was known to occur in some zeolites since the seminal works of Grima²⁷ and Evans.²⁸ However, it was not systematically investigated before the 2015 work by Siddorn *et al.*, which simulated two hundred pure silica zeolites (both at the classical and quantum chemical level) to identify zeolitic frameworks with partial or complete auxeticity.⁹

In this work, we scale up this analysis and use machine learning to speed up the discovery of auxetic zeolite frameworks. In order to investigate the links between structural properties and anomalous mechanical properties in zeolites, we used a database of hypothetical zeolitic

structures generated by Pophale *et al.*²⁹ This database contains 590,811 structures, along with the associated mechanical properties computed at the force field level, that have never been analyzed before. By systematic exploration of this database, we provide a statistical analysis of the mechanical behavior of zeolitic frameworks. We then carried out density functional theory (DFT) calculations of the elastic properties on a specific subset extracted from the database, to unravel the mechanism behind complete auxeticity in all-silica zeolites. Finally, we trained a machine learning algorithm on this DFT data, developing a methodology allowing us to screen such materials for potential auxeticity, efficiently predicting their Poisson’s ratio at small computational cost.

Methods and computational details

Electronic structure and mechanical properties calculations

We used the CRYSTAL14 software in order to perform periodic DFT calculations, with symmetry adaptation, based on localized atomic orbitals.³⁰ We described both O and Si atoms with triple- ζ valence polarized (TZVP) basis sets,³¹ and used a Generalised Gradient Approximation (GGA) exchange–correlation functional, namely the PBE exchange–correlation functional adapted for solids (PBESOL).³² In order to correct for the underestimation the long-range dispersion interactions in DFT at the GGA level, we used dispersion corrections in the Grimme “D2” method.³³ We have previously validated this methodology on the calculation of mechanical properties of both zeolites^{6,34} and metal–organic frameworks.³⁵

Initial structures were taken from the database of Pophale *et al.*²⁹ In that work, they were geometry-optimized using a classical force field, the van Beest–Kramer–van Santen (BKS) potential.³⁶ These were considered as starting points for DFT geometry optimization, with CRYSTAL14 using a quasi-Newton scheme and the Broyden-Fletcher-Goldfarb-Shanno algorithm to update the Hessian at each step. Tighter convergence criteria (by a factor of ten) on both forces and displacements were used to ensure high accuracy in the calculation

of energy derivatives.

Once the structures were relaxed, we calculated their mechanical properties calculations (CRYSTAL14 keyword `ELASTCON`) in the elastic regime, through series of deformations (strains) of the unit cell along each of the 6 deformation modes, and subsequent atomic relaxations.^{37,38} The output of these calculations is the tensor of second-order elastic constants, which is represented in Voigt notation as a 6×6 matrix of elastic constants C_{ij} ,³⁹ where

$$C_{ij} = \frac{1}{V} \left(\frac{\partial^2 E}{\partial \varepsilon_i \partial \varepsilon_j} \right) \quad (1)$$

For highly-symmetric crystals, fewer deformation modes are needed. Such calculations are generally more costly than geometry optimizations, and their computation time depends on a number of factors: the number of deformation modes to be performed; the point group symmetries remaining after individual deformations; and the number of steps needed to reach convergence, after each cell deformation.

In order to visualize and analyze the elastic properties, we used the ELATE program, which is both an open source Python module (available at <https://github.com/fxcoudert/elate>) for the manipulation of elastic tensors and a standalone online application (available at <http://progs.coudert.name/elate>) for the routine analysis of elastic tensors. ELATE allowed us to calculate: i) average mechanical properties in the 3 usual averaging schemes (Voigt, Reuss, and Hill); ii) the eigenvalues of the elastic tensor (including softest and stiffest modes); iii) minima and maxima of the elastic moduli with associated axes; iv) 2D and 3D graphs of the spatial variations of all moduli.

Representative input files for the different DFT calculations described above are available online on our group repository at <https://github.com/fxcoudert/citable-data>

Supervised learning

The main principle of supervised learning is to create a predictor by training an algorithm on two sets of variables in the training dataset: some simple geometric descriptors of each material, obtained at low computational cost, and the required property that one wants to predict (calculated at a higher level of theory). Here we decided to train machine learning algorithms, as implemented in the `scikit-learn` python library,⁴⁰ to create three predictors based on regression methods, targeting the average, minimum and maximum values of the directional Poisson’s ratio. Previous studies harnessing machine learning to predict various properties of crystalline materials have involved a wide range of descriptors: structural and geometric quantities, partial radial distribution functions, energetic properties, etc.^{22,23,41} Following the study of Evans and Coudert,²³ we focus in the present work on the local and structural features of the zeolite structures, summarized in Figure 1. The descriptors used can be classified into three different categories: local descriptors, concerning bond length, angle and dihedrals; global descriptors, such as density or ring sizes distribution; and descriptors related to porosity, like accessible surface area, largest included sphere diameter or accessible volume. The full list of descriptors is given in Table 1. These descriptors have been shown to accurately predict bulk and shear moduli of zeolites with impressive accuracy.²³ Local and global information are obtained from zeolitic structures in CIF format using `pymatgen` library,⁴² while geometric parameters related to the porosity have been computed thanks to `Zeo++` software.^{43,44}

To create the predictor, we used a gradient boosting regressor (GBR)^{45,46} as implemented in Python `scikit-learn` package. This method trains regression trees as an additive model in stepwise approach by optimizing arbitrary loss functions. At each stage, a regression tree is fitted on the negative gradient of the loss function. GBR is both an accurate and effective method that has been used in diverse applications, such as web search ranking.⁴⁷ In this work, we chose this method over others (such as support vector machines⁴⁸) because GBR models are considered robust, interpretable, and applicable for the relatively small data set

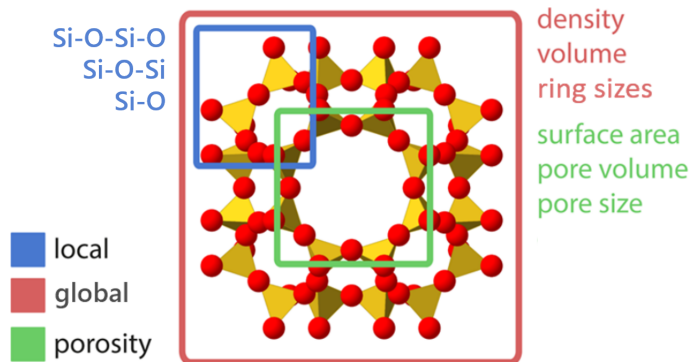


Figure 1: Summary of the descriptors used as entries in the machine learning process classified in local properties, global properties and porosity-related properties.

we study here.⁴⁹ The implementation of GBR employs well-established selection criteria with 3-fold cross-validation, which was repeated 100 times to give a representation of the model accuracy. The hyperparameters were chosen so as to provide good prediction accuracy and minimize overfitting. In particular, as in Ref. 23, we used 1000 estimators, a learning rate of 0.01, a minimum samples split of 2, a minimum samples per leaf of 3, a maximum depth of 3, a maximum number of features equal to the square root of the number of total features and a subsample parameter of 0.4.

Systems studied

As stated before, we focused on the present work on the Pophale et al.²⁹ database for zeolite-like materials. In this database are reported the mechanical properties of 590 811 structures, of which 134 correspondent to experimentally known zeolite structures. Both the structures (resulting from energy minimization) and the elastic properties were calculated at the classical level, using the van Beest–Kramer–van Santen (BKS) force field.³⁶ Starting from the elastic tensors (expressed as a 6×6 symmetric matrix of 21 independent elastic constants C_{ij} .⁵⁰) of all structures in the database, we carried out a first analysis with the ELATE code, focussing on the prevalence of mechanical stability and anomalous mechanical properties. This analysis of the database shows first that the database contains 128 563

Table 1: Full list of descriptors used in the present work.

	descriptor
local	Si-O average/median/standard deviation (Å) Si-O-Si average/median/standard deviation (°) Si-O-Si-O average/median/standard deviation (°) Si-O geometric mean (Å) Si-O-Si geometric mean (°) Si-O harmonic mean (Å) Si-O-Si harmonic mean (°) Si-O skewness Si-O-Si skewness Si-O-Si-O skewness Si-O maximum (Å) Si-O-Si maximum (°) Si-O-Si-O maximum (°) Si-O minimum (Å) Si-O-Si minimum (°) Si-O-Si-O minimum (°)
global	density numbers of N -member rings ($3 \leq N \leq 20$)
porosity	Largest included sphere (Å) Largest free sphere (Å) Largest included free sphere (Å) Accessible surface area for probe radius 1.2 Å (Å ² /Å ³) Non-accessible surface area for probe radius 1.2 Å (Å ² /Å ³) Accessible volume (%) Non-accessible volume (%)

mechanically unstable structures ($\simeq 22\%$ of the database). These structures correspond to local minima in the potential energy surface, but the minima are very shallow and become unstable when perturbed by a small finite strain ($\simeq 1\%$). This highlights the generic need for curation of materials databases, but also demonstrates potential limitations of the BKS force field for the prediction of mechanical properties.

A second finding is that 578 structures are predicted to be completely auxetic, at the force field level; i.e., they have a Poisson's ratio that is negative in all directions of space. That number represents $\simeq 0.1\%$ of the mechanically stable structures. From there, we then conducted DFT calculations on all of these 578 candidate structures for complete auxeticity, in order to check whether the predictions could hold true. In order to have a point of com-

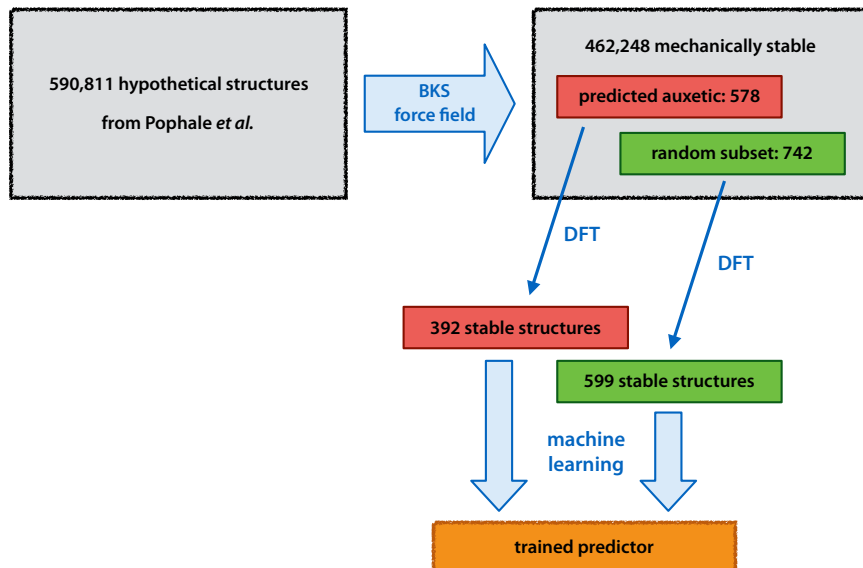


Figure 2: Description of the modelling strategy and different zeolites subsets used in the present work.

parison, and to understand the contrast between auxetic and nonauxetic structures, we took another set of 742 structures, randomly chosen among the rest of the database. After calculations at the DFT level, we restricted ourselves to the structures who DFT calculations converged in the time available, and yielded positive definite second-order elastic tensors corresponding to a mechanically stable structure. From the initial set of 578 candidate auxetic zeolites, we obtained 392 mechanically stable structures; from the random subset of 742 frameworks, we obtained 599 mechanically stable structures. The different stages of our modelling strategy are summarized in Figure 2.

Results and discussion

Elastic properties at the force field level

In this section we analyze the elastic properties at the classical level in the hypothetical zeolites database, computed with BKS force field. We report the noticeable correlations observed throughout the database of mechanical properties, and compare the behavior of the

590,677 hypothetical structures compared to the known zeolitic frameworks (134 structures). Although the database is public, such an analysis was never reported before. These trends obtained at the classical level can also be compared with the previously published DFT results computed in our group for 121 known all-silica zeolites extracted from International Zeolite Association (IZA) Database.⁶

First, energetic stability of zeolitic frameworks in the Pophale database is plotted as a heatmap in density/energy space, in Figure 3. Due to the very large number of structures, a heat map is required to show the density of points in each region, where otherwise only the contour of the populated region would be visible. On top of the heat map, the points corresponding to experimentally known structures are highlighted in red; a linear fit of these know structures is also plotted (red line), indicating the general trend. As previously reported in zeolites and other porous materials with polymorphism, there is a clear negative correlation between energy and density, meaning that the structures get more and more thermodynamically stable as density increases. These calculations at force field level are in line with what was already shown by DFT calculations on a set of 121 known structures presented by Coudert in 2013.⁶ Furthermore, the hypothetical structures seem to follow a similar, although not as clear, trend. We also see that experimental structures seem to favour the low energy region of the plot, although this is not a systematic effect.

Although the force field calculations are able to capture the correlation between energy and density, they fail to reproduce the correlation previously observed by DFT calculations between energetic stability and low elastic anisotropy.⁶ The elastic anisotropy η is defined as a function of the direction Young’s modulus E and shear modulus G :

$$\eta = \max \left(\frac{E_{\max}}{E_{\min}}, \frac{G_{\max}}{G_{\min}} \right) \quad (2)$$

It was observed that experimentally known zeolites have relatively low elastic anisotropy, and this has been proposed as a factor to determine the experimental feasibility of hypothetical

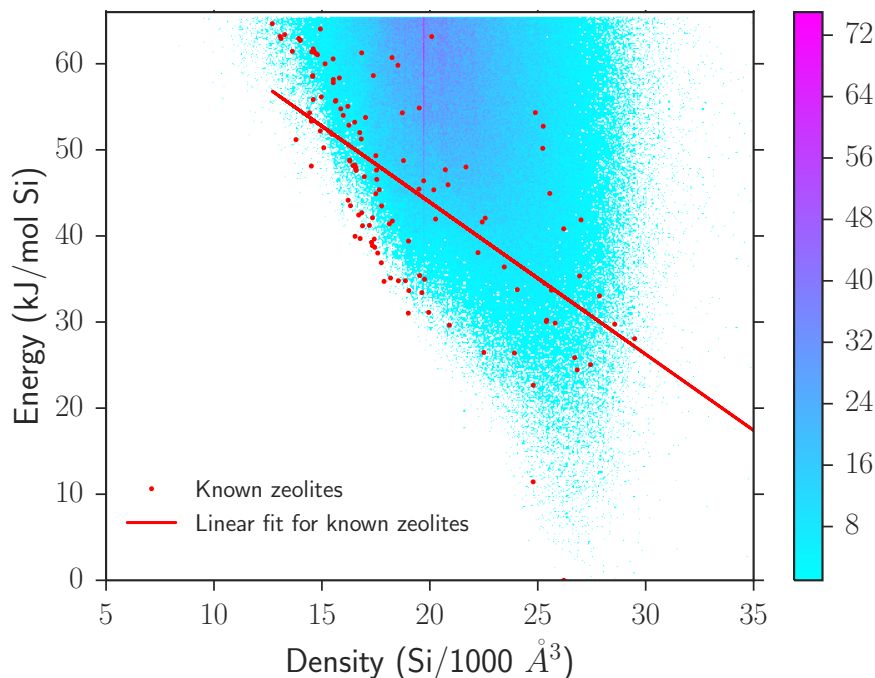


Figure 3: Correlation between the energy relative to α -quartz (in kJ/mol) and the density (in Si atoms per 1000 \AA^3), plotted as a heat map. Red points indicate know experimental zeolitic structures. The red line is a linear fit of those points.

structures. The physical mechanism behind this is that high elastic anisotropy indicates a limited mechanical stability of the material.⁶

Figure 4 shows the very weak correlation (again as a heat map) between elastic anisotropy and energy. We see that at the classical level, a majority of structures have low elastic anisotropy ($\eta \leq 5$), both for hypothetical and experimental frameworks. However, unlike what has been demonstrated at the DFT level, there is also a minority of known materials with high elastic anisotropy ($\eta > 5$). Moreover, there are structures with both high energy (more than 20 kJ/mol above α -quartz) and high anisotropy, whereas DFT calculations showed that an elastic anisotropy under 4 meant a very high probability of having an energy under 20 kJ/mol.⁶ The preference for lower anisotropy is thus reproduced to a limited extent, and the correlation between energetic stability and elastic anisotropy is not clear at the force field level. This is an important limitation, as mechanical anisotropy has consequences on the macroscopic behavior and stability of materials.^{51,52}

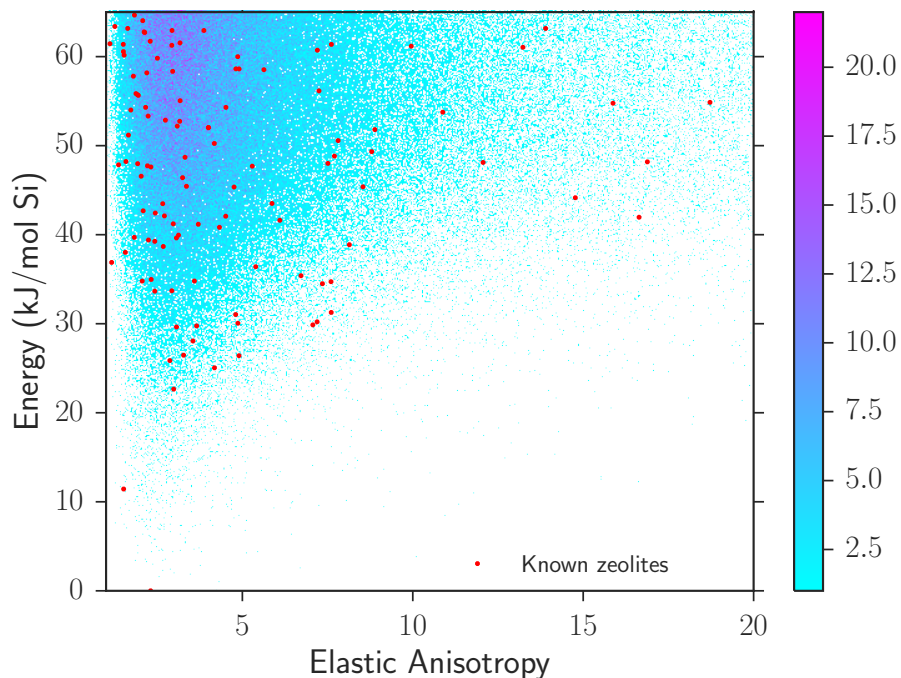


Figure 4: Correlation between the elastic anisotropy (unitless) and the framework energy relative to α -quartz (in kJ/mol), shown as a heat map. Red points indicate know experimental zeolitic structures.

In a similar vein, a remarkable positive correlation between the minimum shear modulus and minimum Young’s modulus for known zeolites was pointed out in Ref. 6. We find here that this behavior is well captured with force field calculations (see right panel of Figure 5), as both known and hypothetical structures lie close a straight line in (G_{\min}, E_{\min}) space. At this stage, this behaviour appears to be a generic property of any zeolitic structure, since the global heat map follows the trends for known zeolites with similar deviations. Furthermore, the left panel of Figure 5 shows that this correlation is even clearer when looking at the average values of the directional shear and Young’s moduli ($\langle G \rangle$ and $\langle E \rangle$, respectively). It means that the basic SiO_4 tetrahedral unit, from which zeolitic frameworks are built, creates a strong coupling between the response to longitudinal and shear stresses — and this, no matter the way the tetrahedra are assembled to form the crystal structure. These findings reinforce the conclusions drawn earlier (ref. 6), this time on a much larger database of zeolitic structures (both experimental and hypothetical).

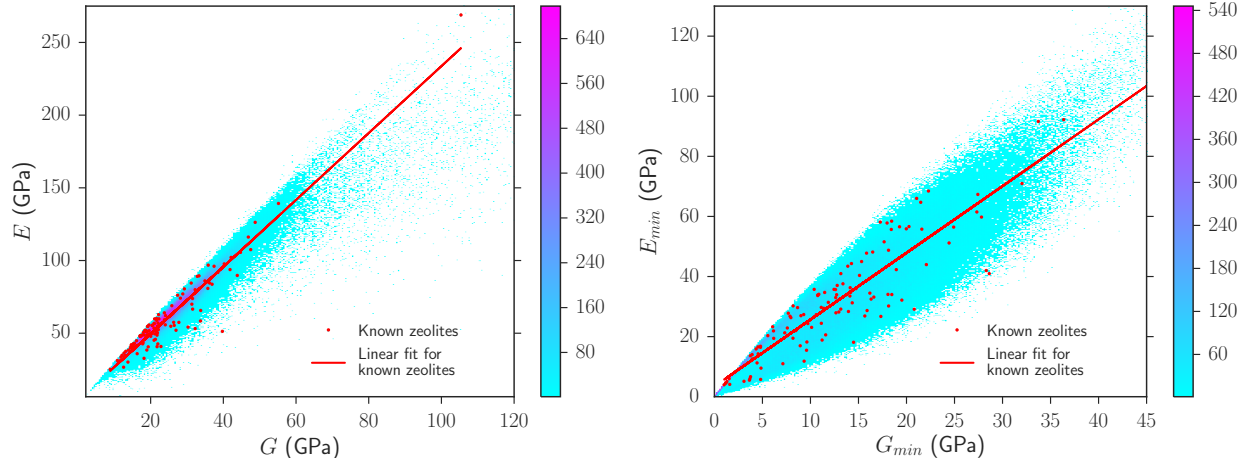


Figure 5: Correlation heatmap between the average shear modulus $\langle G \rangle$ (GPa) and the average Young's modulus $\langle E \rangle$ (GPa) (left). Correlation heatmap between the minimum shear modulus G_{\min} (GPa) and the minimum Young's modulus E_{\min} (GPa) (right).

If we now focus specifically on the subset of 578 auxetic structures present in Pophale database, we show in the supporting information a zoom of Figure 5 where those materials are in green. The positive correlation between shear and Young's moduli appears clearly for this subset too, but with lower Young's moduli overall. Another interesting feature is the average relative energy of the auxetic subset, which is very close to the one for all the zeolites, only 3 kJ/mol higher than the average for known structures. These first characteristics suggest that the auxetic structures identified should be, for some at least, experimentally feasible.

Finally, we turn our attention to the correlation between extreme values of the Poisson's ratio and elastic anisotropy. It was shown in studies at the DFT level (on a small number of frameworks) that known zeolites behave similarly to dense silica crystalline polymorphs in terms of minimum and maximum Poisson's ratio: namely, that they follow two separate families of curves when plotted against Ledbetter anisotropy.⁹ Ledbetter anisotropy is defined as the square of the maximum shear-sound-wave velocity divided by the square of the minimum shear-sound-wave velocity⁵³ — similarly to the elastic anisotropy η defined above, it characterizes the mechanical anisotropy of a crystal. In Figure 6, we plotted the minimum and maximum Poisson's ratio for all the zeolites in the database, highlighting the known structures. The dual evolution of the extremal Poisson's ratio with respect to elastic anisotropy is

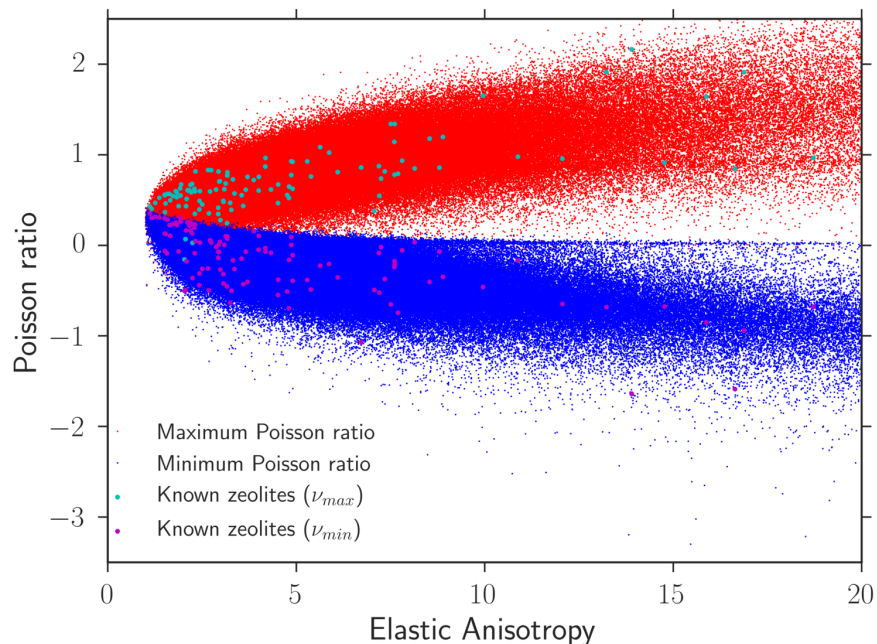


Figure 6: Correlation between the elastic anisotropy η , and the minimum and maximum Poisson’s ratio (unitless). Maximum Poisson’s ratio is in red, minimum Poisson’s ratio in blue. Values for known zeolites are highlighted in cyan and pink, respectively.

found to be reproduced at a much larger scale than ever probed before. Moreover, it occurs for both known zeolites as well as hypothetical structures, showing a generic trend across the entire family of materials.

How accurate are force field predictions?

Because of the correlations between different mechanical and energetic properties observed above, at the large scale on the hypothetical zeolites database using force field calculations, we wanted to compare force field-level properties with computationally expensive DFT results, obtained on two subsets of the database: 392 and 599 structures for the completely auxetic and random subsets, respectively.

Structural properties

As exemplified in Figure 7 with lattice parameters a and α (but the trend is the same for b , c , β and γ), the cell parameters computed at the force field level are statistically in good

agreement with the ones computed at the DFT level. There are a few outliers, indicating frameworks that have a significantly different structure after force field or DFT geometry optimization. One example is materials #9316681 (numbered in the database from Pophale et al²⁹), for which BKS predicted $a = 12.4 \text{ \AA}$ and DFT gives $a = 15.8 \text{ \AA}$ (i.e. an increase of 27%). Statistically, the volumes are also in good agreement between DFT and force field calculations. Furthermore, the agreement is similar for both subsets characterized.

The agreement between DFT and the BKS force field shows larger fluctuations, shown in Figure 8. The overall agreement is still correct, yet we see here a different effect between the two subsets: the maximum deviation observed is 12% for the random subset, but goes up to 35% for one of the structures in the auxetic subset (#9420191). The results obtained on these two sets with several hundreds of structures show that force field based calculations of structural properties are suitable as a first stage in a high-throughput screening strategy for novel zeolitic structures, with an average deviation well below 10%. However, we see that this performance is not homogeneous across the range of materials: frameworks with some “unusual” features, such as negative Poisson’s ratio, may be predicted with lower accuracy.

Energy and mechanical properties

Turning now to the energetics of the zeolitic frameworks, we show in Figure 9 the energies (relative to α -quartz) given by the DFT calculations compared to the results obtained by force field calculations. We first see that the results, both for the auxetic candidates and random subsets, have a systematic difference. There is a clear trend for the BKS force field to overestimate the energies (the root mean square error and the mean absolute error are both above 26 kJ/mol for both subsets). But beyond this systematic effect, there is a wide spread of values themselves (correlations of $r = 0.46$ for the auxetic subset and $r = 0.38$ for the random subset). More importantly, we see a nonuniform distribution of energies in the case of auxetic candidate structures (Figure 9, left panel) which indicates that the BKS results are missing some important features of the relative energies. This highlights the limitations

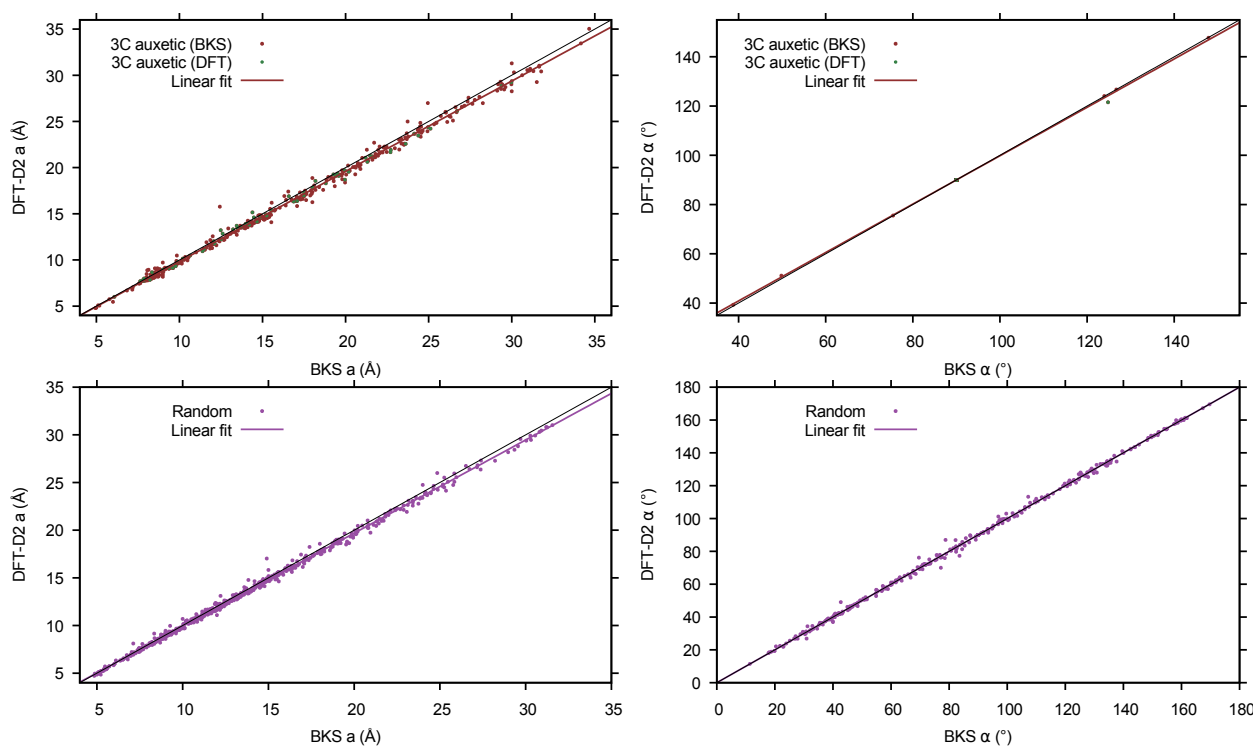


Figure 7: Comparison between unit cell parameters a (left) and α (right) obtained by DFT calculations and values given by the BKS force field. Top panel: subset of structures predicted to be auxetic in the BKS calculations. Bottom panel: random subset of hypothetical zeolites. The $y = x$ line is indicated in black in all plots.

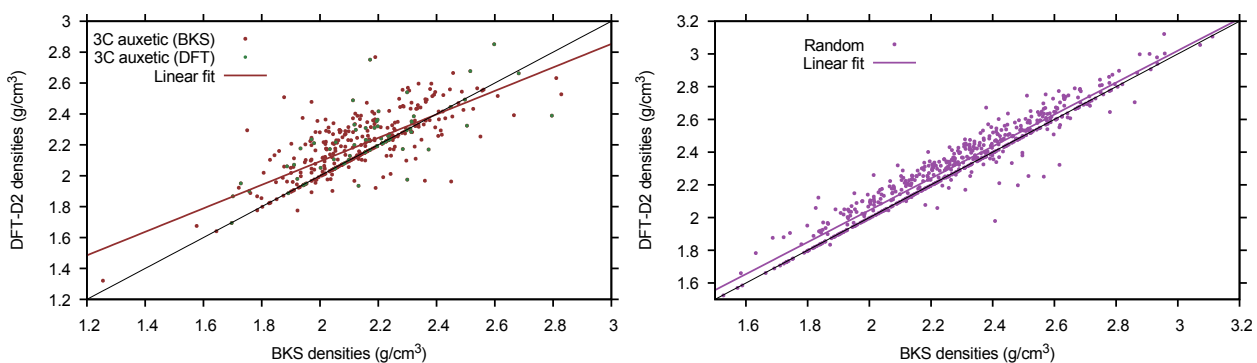


Figure 8: Comparison between the density obtained in DFT calculations and the values given by the BKS force field. Left: subset of structures predicted to be auxetic in the BKS calculations. Right: random subset of hypothetical zeolites. The $y = x$ line is indicated in black in both plots.

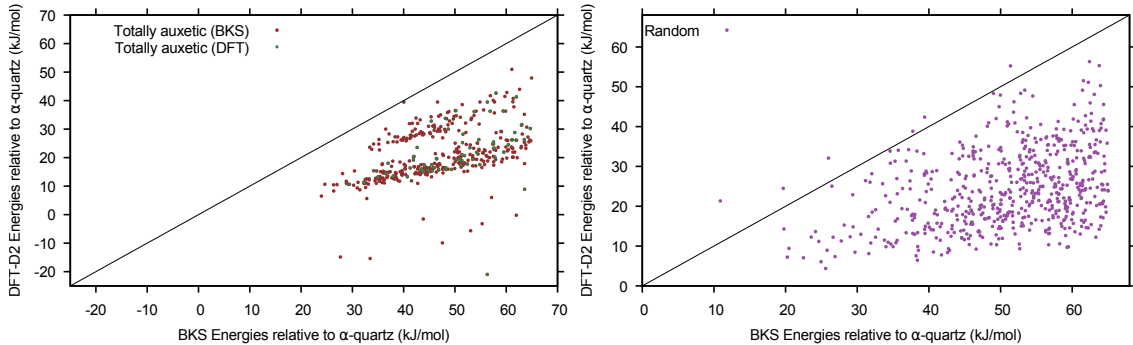


Figure 9: Comparison of relative energies obtained by DFT calculations, compared to the BKS force field. Left: for the subset of candidate auxetic structure; right: for the random subset of frameworks.

of the use of the BKS force fields (despite being very standard in silica polymorphs atomistic simulations) for the prediction of energetic features of frameworks.

Given the poor results of the BKS force field for energy prediction, it was important to test its validity for mechanical properties; indeed, the second-order elastic tensor is a second derivative of the energy with respect to unit cell strain. The performance of the force field is therefore found to be quite bad: Table 2 shows three estimators of the deviations between DFT values and force field values. Even taking a quite lenient measure, the relative root mean square error (i.e., root mean square error divided by the range of the property in the subset considered), the lowest value is attained for the bulk modulus K around 20%. That amounts to a mean absolute error for the auxetic subset of 7.9 GPa, which represents almost half of the average value of the DFT bulk modulus for the auxetic subset (16 GPa) and more than the standard deviation over this subset of 392 structures (6.5 GPa). The average Young’s modulus E is also found to be in the same range of error as the bulk modulus, while the average shear modulus G predictions are even further off.

As our ultimate goal here is to determine whether a material has some degree of auxeticity, we were particularly interested in the the extremal values of the Poisson’s ratio, ν_{\min} and ν_{\max} . Table 2 clearly shows how unreliable the force field data are in that regard. In fact, the correlation between the force field and the DFT values is simply nonexistent, and the very large values for the RMSE show that outliers are very badly described. The large values

Table 2: Relative root mean square error (rRMSE, RMSE divided by the range of the property considered), the mean absolute error (MAE) and the Pearson’s correlation coefficient (r) for the energy (E_n), the bulk modulus (K), the Young’s modulus (E), the shear modulus (G), the Poisson’s ratio (ν) and the minimum and maximum of the Poisson’s ratio (ν_{\min} and ν_{\max}) for auxetic and random subsets.

	E_n (kJ/mol)	K (GPa)	E (GPa)	G (GPa)	ν	ν_{\min}	ν_{\max}
rRMSE (auxetic)	39%	22%	19%	174%	47%	1.93	1.5
rRMSE (random)	55%	21%	85%	78%	46%	1.35	13
MAE (auxetic)	26	7.9	11	10	0.29	0.71	0.53
MAE (random)	27	10	19	7.6	0.07	0.45	3.4
r (auxetic)	0.46	0.29	0.43	-0.01	0.002	0.07	0.04
r (random)	0.38	0.60	0.25	0.22	0.21	0.03	0.09

of the mean absolute error also shows that the force field is unable to be accurate even for “normal” materials (outside the auxetic group). This demonstrates, for the first time, clearly the impossibility of using the state-of-the-art BKS force field for large-scale screening of auxetic zeolites. We have therefore turned to the machine learning techniques, using the DFT results as training data set, to allow for rapid prediction of auxeticity in these zeolitic structures.

Machine learning for predicting auxeticity

Given the lackluster performance of force field-based calculations to predict the mechanical properties of zeolitic frameworks, and in particular the anisotropic properties such as the Poisson’s ratio, we have sought an alternative approach. Because DFT is too computational intensive to serve as a basis for high-throughput screening of large databases of real or hypothetical structures, we turned to a machine learning approach to predict the value of Poisson’s ratio for arbitrary zeolitic structures. The details of the supervised learning (training and the cross-validation) are given in the Methods section. This work is a significant extension to the work previously performed.²³ That earlier study, on a much smaller data set (121 zeolites instead of 991 structures here), demonstrated the capability of machine learning to predict *average* mechanical properties — namely, the bulk and shear moduli. We focus

here on the more difficult task of the Poisson’s ratio.

The results are represented in Figure 10, and Table 3 gives the detail of the relative errors of the model — broken down between the random and auxetic subsets of structures. Overall, the results given by the GBR predictor are much better than the ones given by force field calculations. If we take, for example, the root mean square error (RMSE) as a measure of the accuracy of each method, compared to the DFT values, it is 0.29 in the case of BKS, while it goes down to 0.12 for the machine learning results. It is also interesting to look at the performance of the methods on each subset: auxetic candidates, random structures, and the known zeolitic frameworks. We can see that the BKS force field has an accuracy almost as good as the GBR predictor for the known structures; this can be explained by the fact that BKS was optimized on known silica polymorphs, and therefore performs better there. However, for more “unusual” structures, and especially for the auxetic subset, the machine learning algorithm has much better performance than the force field.

In addition to the prediction of the mechanical properties themselves, one of the nice feature of the GBR model is that it is not a “black box” model, but it does also provide as output, after training the algorithm, the relative importance of the different descriptors provided as input. One can therefore produce a second generation of algorithm, either by introducing new descriptors, or by removing those that had low weight in the original results. Figure 11 presents the relative importance of the descriptors with a relative importance above 40% (compared to the most important one). It is interesting to see that local descriptors (see Figure 1 and Table 1), and especially the amplitudes of their variations are most represented with 8 descriptors over 16. In terms of porosity-related descriptors, the linear sizes and the relative surface areas matter more than the actual porous volumes. Among the global descriptors, the topology seem to play a big role with specific ring sizes which have similar importance, namely the proportion of 4, 5, 6 and 8-member rings.

Finally, having found that the extremal values of the directional Poisson’s ratio (minimum and maximum) were particularly badly predicted at the force field level (Table 2), we applied

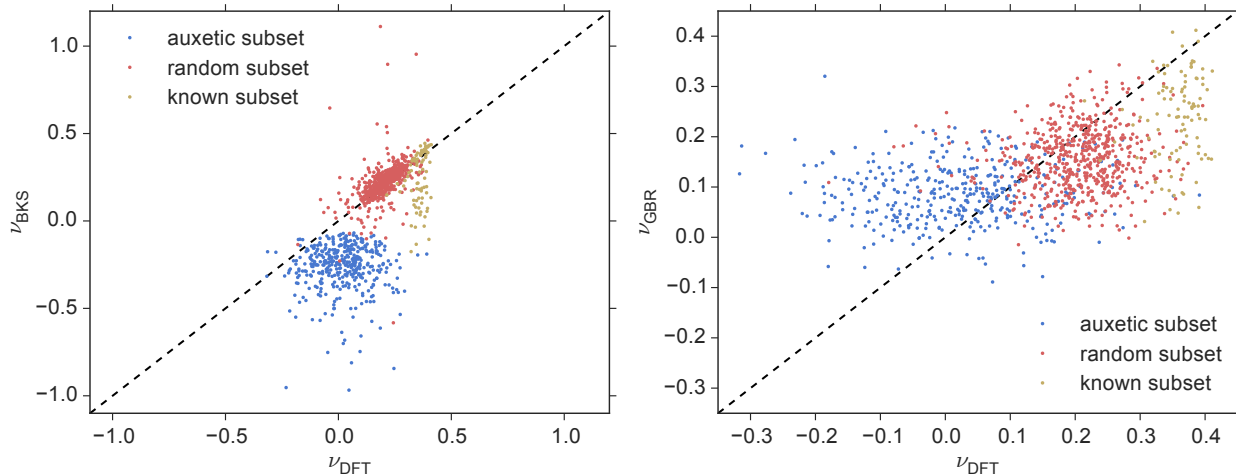


Figure 10: Average Poisson’s ratio values obtained by the BKS force field (left panel) and the machine learning algorithm (right panel), compared in each case against the results of DFT calculations. Note that the scales are different in both graphs, due to the smaller overall dispersion obtained with machine learning. Structures from the auxetic candidates subset are highlighted in blue, from the random subset in red and from the known zeolites in green.

Table 3: Root mean square error (RMSE) and mean absolute error (MAE) for different subsets of structures (and the total data set) in the prediction of the Poisson’s ratio.

Subset (method)	RMSE	MAE
all (GBR)	0.12	0.096
all (BKS)	0.29	0.15
auxetic subset (GBR)	0.15	0.12
auxetic subset (BKS)	0.34	0.29
random subset (GBR)	0.10	0.079
random subset (BKS)	0.26	0.070
known zeolites (GBR)	0.15	0.13
known zeolites (BKS)	0.18	0.12

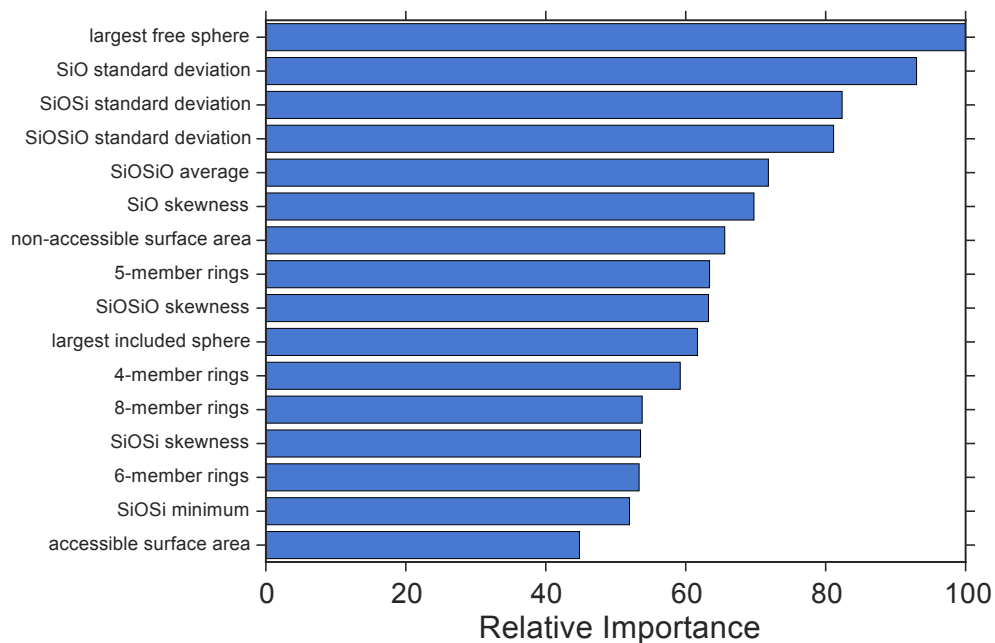


Figure 11: Relative importance of the most important descriptors in the GBR model.

the same methodology to ν_{\min} and ν_{\max} , to check whether the GBR model would be able to be trained into an accurate predictor for this particularly difficult problem. It turns out to be the case, as shown in Table 4. Although the errors are relatively important (more than for the average Poisson’s ratio), they are much lower for the GBR predictor than for the results obtained with the BKS force field. In particular, looking at the errors for all the structures, the GBR model provides errors below 0.5, even for the RMSE which emphasizes the extreme cases. These are very promising results, highlighting a relatively low-cost model for the prediction and high-throughput screening of auxeticity in zeolitic frameworks.

Table 4: Root mean square error (RMSE) and mean absolute error (MAE) for the 3 subsets and their assembly for the prediction of the Poisson’s ratio.

Subset (method)	RMSE (ν_{\min})	MAE (ν_{\min})	RMSE (ν_{\max})	MAE (ν_{\max})
all (GBR)	0.39	0.26	0.46	0.32
all (BKS)	1.4	0.51	9.8	2.1
auxetic subset (GBR)	0.38	0.25	0.46	0.34
auxetic subset (BKS)	1.5	0.66	0.62	0.47
random subset (GBR)	0.40	0.25	0.48	0.31
random subset (BKS)	1.4	0.45	13	3.4
known zeolites (GBR)	0.43	0.35	0.34	0.27
known zeolites (BKS)	0.50	0.35	0.47	0.32

Conclusions and perspectives

In this work, we have explored the mechanical properties of zeolitic frameworks, both experimentally known and hypothetical, at different levels of theory. By analyzing the behavior of about 600,000 hypothetical zeolitic structures at the force field level (classical simulations), we have shown that the correlations between various mechanical properties, found previously on a much smaller number of materials, are very general. Moreover, we showed that the prediction of mechanical properties at the force field level is widely inaccurate for anisotropic properties, such as the Poisson’s ratio. This was done by performing DFT calculations on a subset of 991 zeolitic frameworks, the largest every systematically investigated at this level (previous works studied at most 121 zeolites). These results showed that force field calculations cannot be used to reliably screen zeolites databases for mechanical metamaterials, such as auxetic materials.

Therefore, we used the DFT data on these 991 zeolitic frameworks as a training set for a machine learning algorithm. While this approach for the prediction of mechanical properties was previously only demonstrated on average properties such as bulk and shear moduli, we show here that machine learning offers a way to build fast and reliable predictive models for anisotropic properties as well, and to predict the Poisson’s ratio and its extremal values. The accuracies obtained are, in particular, much better than the current “cheap” approach for screening, which is the use of force fields. These results are a significant improvement over the previous work, due to the more difficult nature of the properties studied, namely the anisotropic elastic response. It is also the first time such a large training data set is used for zeolitic materials. This study paves the way to broader use of the machine learning techniques for the calculation of mechanical properties at low computational cost, and its use in high-throughput screening methodologies.

Finally, while the present study focused on a family of materials with identical chemical composition (SiO_2 polymorphs), it can be applied to other classes of materials. We will be working in the future to extend the methodology to materials with different chemical

composition, by adding descriptors based on chemical composition, provided that we can exploit or generate large reliable databases, based on either experimental data or quantum calculations. It will be interesting, in particular, to see how the machine learning approach used here can be transferred to other classes of materials with identical four-connected net topologies, such as carbon allotropes, zinc cyanide polymorphs, zeolitic imidazolate frameworks, etc. This will require the use of descriptors describing the chemical composition of the materials,⁵⁴ or the properties of the bonds involved. Further down the line, the methodology could be extended to non-siliceous zeolites, with descriptors related to the chemical nature of the tetrahedral atoms, compensating cations, and Si/Al ratio.

Acknowledgement

We acknowledge access to high-performance computing platforms provided by GENCI grant A0070807069. This work benefited from the financial support of ANRT (thèse CIFRE 2015/0268). We thank Michael Deem for discussion about his hypothetical zeolite databases, and the 2013 Gordon Research Conference on “Nanoporous Materials and Their Applications” for starting this project.

Supporting Information Available

This material is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- (1) Treacy, M.; Randall, K.; Rao, S.; Perry, J.; Chadi, D. Enumeration of periodic tetrahedral frameworks. *Z. Kristall.* **1997**, *212*, 768–791.
- (2) *Database of Zeolite Structures*, IZA Structure Commission, available online at <http://www.iza-structure.org/databases/>.

- (3) Woodley, S. M.; Catlow, R. Crystal structure prediction from first principles. *Nature Mater.* **2008**, *7*, 937–946.
- (4) Friedrichs, O. D.; Dress, A. W. M.; Huson, D. H.; Klinowski, J.; Mackay, A. L. Systematic enumeration of crystalline networks. *Nature* **1999**, *400*, 644–647.
- (5) Treacy, M.; Rivin, I.; Balkovsky, E.; Randall, K.; Foster, M. Enumeration of periodic tetrahedral frameworks. II. Polynodal graphs. *Micropor. Mesopor. Mat.* **2004**, *74*, 121–132.
- (6) Coudert, F.-X. Systematic investigation of the mechanical properties of pure silica zeolites: stiffness, anisotropy, and negative linear compressibility. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16012.
- (7) Astala, R.; Auerbach, S. M.; Monson, P. A. Density Functional Theory Study of Silica Zeolite Structures: Stabilities and Mechanical Properties of SOD, LTA, CHA, MOR, and MFI. *J. Phys. Chem. B* **2004**, *108*, 9208–9215.
- (8) Li, Z.; Johnson, M. C.; Sun, M.; Ryan, E. T.; Earl, D. J.; Maichen, W.; Martin, J. I.; Li, S.; Lew, C. M.; Wang, J.; Deem, M. W.; Davis, M. E.; Yan, Y. Mechanical and Dielectric Properties of Pure-Silica-Zeolite Low-k Materials. *Angew. Chem. Int. Ed.* **2006**, *45*, 6329–6332.
- (9) Siddorn, M.; Coudert, F.-X.; Evans, K. E.; Marmier, A. A systematic typology for negative Poisson’s ratio materials and the prediction of complete auxeticity in pure silica zeolite JST. *Phys. Chem. Chem. Phys.* **2015**, *17*, 17927–17933.
- (10) Coudert, F.-X.; Evans, J. D. Nanoscale metamaterials: Meta-MOFs and framework materials with anomalous behavior. *Coord. Chem. Rev.* **2019**, *388*, 48–62.
- (11) Grima-Cornish, J. N.; Grima, J. N.; Attard, D. Negative Mechanical Materials and

- Metamaterials: Giant Out-of-Plane Auxeticity from Multi-Dimensional Wine-Rack-like Motifs. *MRS Advances* **2020**, 1–9.
- (12) Bouëssel du Bourg, L.; Ortiz, A. U.; Boutin, A.; Coudert, F.-X. Thermal and mechanical stability of zeolitic imidazolate frameworks polymorphs. *APL Mater.* **2014**, *2*, 124110.
- (13) de Jong, M.; Chen, W.; Angsten, T.; Jain, A.; Notestine, R.; Gamst, A.; Sluiter, M.; Ande, C. K.; van der Zwaag, S.; Plata, J. J.; Toher, C.; Curtarolo, S.; Ceder, G.; Persson, K. A.; Asta, M. Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2015**, *2*, 150009.
- (14) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501–1509.
- (15) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **2013**, *3*, 2810.
- (16) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29*, 2844–2854.
- (17) Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. What Are the Best Materials To Separate a Xenon/Krypton Mixture? *Chem. Mater.* **2015**, *27*, 4459–4475.
- (18) Calfa, B. A.; Kitchin, J. R. Property prediction of crystalline solids from composition and crystal structure. *AIChE Journal* **2016**, *62*, 2605–2613.
- (19) de Albuquerque, V. H. C.; Cortez, P. C.; de Alexandria, A. R.; Tavares, J. M. R. A new solution for automatic microstructures analysis from images based on a backpropagation artificial neural network. *Nondestruct. Test. Evaluation* **2008**, *23*, 273–283.

- (20) Walsh, A., Sokol, A. A., Catlow, C. R. A., Eds. *Computational Approaches to Energy Materials*; John Wiley & Sons Ltd, 2013.
- (21) Kiselyova, N.; Gladun, V.; Vashchenko, N. Computational materials design using artificial intelligence methods. *J. Alloys Compd.* **1998**, *279*, 8–13.
- (22) de Jong, M.; Chen, W.; Notestine, R.; Persson, K.; Ceder, G.; Jain, A.; Asta, M.; Gamst, A. A Statistical Learning Framework for Materials Science: Application to Elastic Moduli of k -nary Inorganic Polycrystalline Compounds. *Sci. Rep.* **2016**, *6*.
- (23) Evans, J. D.; Coudert, F.-X. Predicting the Mechanical Properties of Zeolite Frameworks by Machine Learning. *Chem. Mater.* **2017**, *29*, 7833–7839.
- (24) Cairns, A. B.; Goodwin, A. L. Negative linear compressibility. *Phys. Chem. Chem. Phys.* **2015**, *17*, 20449–20465.
- (25) Chibani, S.; Coudert, F.-X. Systematic exploration of the mechanical properties of 13 621 inorganic compounds. *Chem. Sci.* **2019**, *10*, 8589–8599.
- (26) Brańka, A. C.; Heyes, D. M.; Wojciechowski, K. W. Auxeticity of cubic materials. *Phys. Status Solidi B* **2009**, *246*, 2063–2071.
- (27) Grima, J. N.; Gatt, R.; Zammit, V.; Williams, J. J.; Evans, K. E.; Alderson, A.; Walton, R. I. Natrolite: A zeolite with negative Poisson’s ratios. *J. Appl. Phys.* **2007**, *101*, 086102.
- (28) Sanchez-Valle, C.; Lethbridge, Z. A. D.; Sinogeikin, S. V.; Williams, J. J.; Walton, R. I.; Evans, K. E.; Bass, J. D. Negative Poisson’s ratios in siliceous zeolite MFI-silicalite. *J. Chem. Phys.* **2008**, *128*, 184503.
- (29) Pophale, R.; Cheeseman, P. A.; Deem, M. W. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12407.

- (30) Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D’Arco, P.; Noël, Y.; Causà, M.; Rérat, M.; Kirtman, B. CRYSTAL14: A program for the ab initio investigation of crystalline solids. *Int. J. Quantum Chem.* **2014**, *114*, 1287–1317.
- (31) Peintinger, M. F.; Oliveira, D. V.; Bredow, T. Consistent Gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations. *J. Comput. Chem.* **2013**, *34*, 451–459.
- (32) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2008**, *100*.
- (33) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (34) Evans, J. D.; Fraux, G.; Gaillac, R.; Kohen, D.; Trouselet, F.; Vanson, J.-M.; Coudert, F.-X. Computational Chemistry Methods for Nanoporous Materials. *Chem. Mater.* **2016**, *29*, 199–212.
- (35) Ortiz, A. U.; Boutin, A.; Coudert, F.-X. Prediction of flexibility of metal–organic frameworks CAU-13 and NOTT-300 by first principles molecular simulations. *Chem. Commun.* **2014**, *50*, 5867.
- (36) van Beest, B. W. H.; Kramer, G. J.; van Santen, R. A. Force fields for silicas and aluminophosphates based on ab initio calculations. *Phys. Rev. Lett.* **1990**, *64*, 1955–1958.
- (37) Perger, W.; Criswell, J.; Civalleri, B.; Dovesi, R. Ab-initio calculation of elastic constants of crystalline systems with the CRYSTAL code. *Computer Physics Communications* **2009**, *180*, 1753–1759.

- (38) Erba, A.; Mahmoud, A.; Orlando, R.; Dovesi, R. Elastic properties of six silicate garnet end members from accurate ab initio simulations. *Phys Chem Minerals* **2014**, *41*, 151–160.
- (39) Marmier, A.; Lethbridge, Z. A.; Walton, R. I.; Smith, C. W.; Parker, S. C.; Evans, K. E. ELAM: A computer program for the analysis and representation of anisotropic elastic properties. *Comput. Phys. Commun.* **2010**, *181*, 2102–2115.
- (40) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (41) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K. Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO₂ Capture. *J. Phys. Chem. Lett.* **2014**, *5*, 3056–3060.
- (42) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319.
- (43) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Micro. Meso. Mater.* **2012**, *149*, 134–141.
- (44) Martin, R. L.; Smit, B.; Haranczyk, M. Addressing Challenges of Identifying Geometrically Diverse Sets of Crystalline Porous Materials. *J. Chem. Info. Model.* **2011**, *52*, 308–318.
- (45) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232.

- (46) Friedman, J.; Hastie, T.; Tibshirani, R. *The elements of statistical learning*; Springer series in statistics New York, 2001; Vol. 1.
- (47) Mohan, A.; Chen, Z.; Weinberger, K. Web-Search Ranking with Initialized Gradient Boosted Regression Trees. *JMLR Workshop and Conference Proceedings: Proceedings of the Yahoo! Learning to Rank Challenge* **2011**, *14*, 77–89.
- (48) Li, H.; Liang, Y.; Xu, Q. Support vector machines and its applications in chemistry. *Chemom. Intell. Lab. Syst.* **2009**, *95*, 188–198.
- (49) Caruana, R.; Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. Proceedings of the 23rd international conference on Machine learning ICML. 2006.
- (50) Nye, J. F. *Physical properties of crystals: their representation by tensors and matrices*; Oxford University Press, 1985.
- (51) Tan, J. C.; Bennett, T. D.; Cheetham, A. K. Chemical structure, network topology, and porosity effects on the mechanical properties of Zeolitic Imidazolate Frameworks. *Proc. Nat. Acad. Sci.* **2010**, *107*, 9938–9943.
- (52) Varughese, S.; Kiran, M. S. R. N.; Ramamurty, U.; Desiraju, G. R. Nanoindentation in Crystal Engineering: Quantifying Mechanical Properties of Molecular Crystals. *Angew. Chem. Int. Ed.* **2013**, *52*, 2701–2712.
- (53) Ledbetter, H.; Migliori, A. A general elastic-anisotropy measure. *J. Appl. Phys.* **2006**, *100*, 063516.
- (54) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput Mater* **2016**, *2*, 364.

Graphical TOC Entry

